Homework 1

Varun Saxena

Enrollment Data

1. How many observations exist in your current dataset?

19126783

2. How many different *plan_types* exist in the data?

There are 27 unique plan types that exist in the data.

3. Provide a table of the count of plans under each plan type in each year.

Table 1: Plan types by year

plan_type	2007	2008	2009	2010	2011	2012	2013	2014	2015
1876 Cost	5855	5459	5825	6035	6851	7633	7731	7069	7157
Continuing Care Retirement Community	95	122	158	142	0	0	0	0	0
ESRD I	75	122	123	117	0	0	0	0	0
ESRD II	12	12	7	8	0	0	0	0	0
Employer Direct PFFS	3247	0	0	0	0	0	0	0	0
Employer/Union Only Direct Contract PDP	32358	29113	25860	28700	28697	28669	25526	25528	25630
HCPP - 1833 Cost	13	13	3938	3604	11	11	10	9	9
HMO/HMOPOS	60012	70176	479978	506802	528473	507272	530909	523304	479275
Local PPO	17427	38470	405197	417551	515700	636701	633884	664716	704993
MA Health Senior Care Options	73	0	0	0	0	0	0	0	0
MN Disability Health Options	21	0	0	0	0	0	0	0	0
MN Senior Health Options	968	0	0	0	0	0	0	0	0
MSA	4422	16515	12267	135	6421	6416	6431	6449	6518
MSA Demo	3274	0	0	0	0	0	0	0	0
Medicare Prescription Drug Plan	920058	963478	945794	893609	771694	815223	826907	1122209	991457
National PACE	405	548	616	717	781	858	953	1118	1216
PFFS	364285	630756	683361	385733	45781	36423	31919	24905	13658
PSO (Federal Waiver of State License)	162	0	0	0	0	0	0	0	0
PSO (State License)	421	535	87	123	176	171	0	0	0
Pilot	15	12	201	53	3	3	2	2	2
Regional PPO	26402	27990	25943	24442	22773	21602	19970	19773	17578
SHMO	1125	0	0	0	0	0	0	0	0
WI Partnership Program	42	0	0	0	0	0	0	0	0
Employer/Union Only Direct Contract PFFS	0	3332	3335	3332	3329	3323	0	0	0
RFB PFFS	0	0	3006	0	0	0	0	0	0
NA	0	0	27505	277533	0	0	0	0	0
Medicare-Medicaid Plan HMO/HMOPOS	0	0	0	0	0	0	265	1319	4130

4. Remove all special needs plans (SNP), employer group plans (eghp), and all "800-series" plans. Provide an updated table after making these exclusions.

Table 2: Revised plan types by year

plan_type	2007	2008	2009	2010	2011	2012	2013	2014	2015
Continuing Care Retirement Community	27	56	98	78	0	0	0	0	0
ESRD II	12	12	7	8	0	0	0	0	0
HMO/HMOPOS	8902	15551	14980	12465	11610	11653	12875	12650	13116
Local PPO	907	1396	1622	1242	626	1049	3429	858	833
MA Health Senior Care Options	73	0	0	0	0	0	0	0	0
MN Disability Health Options	21	0	0	0	0	0	0	0	0
MN Senior Health Options	968	0	0	0	0	0	0	0	0
PSO (Federal Waiver of State License)	12	0	0	0	0	0	0	0	0
PSO (State License)	14	44	12	26	35	28	0	0	0
Regional PPO	4729	7406	5918	3718	2325	2132	3563	2430	2378
SHMO	592	0	0	0	0	0	0	0	0
WI Partnership Program	42	0	0	0	0	0	0	0	0

5. Merge the the contract service area data to the enrollment data and restrict the data only to contracts that are approved in their respective counties. Limit your dataset only to plans with non-missing enrollment data. Provide a graph showing the average number of Medicare Advantage enrollees per county from 2008 to 2015.

I used the inner join function to join the relevant datasets. Then, I used a filter to exclude entries with missing data. The resulting plot contains the graph of Medicare Advantage enrollees.

```
Warning: Removed 21405 rows containing non-finite values (`stat_summary()`).
```

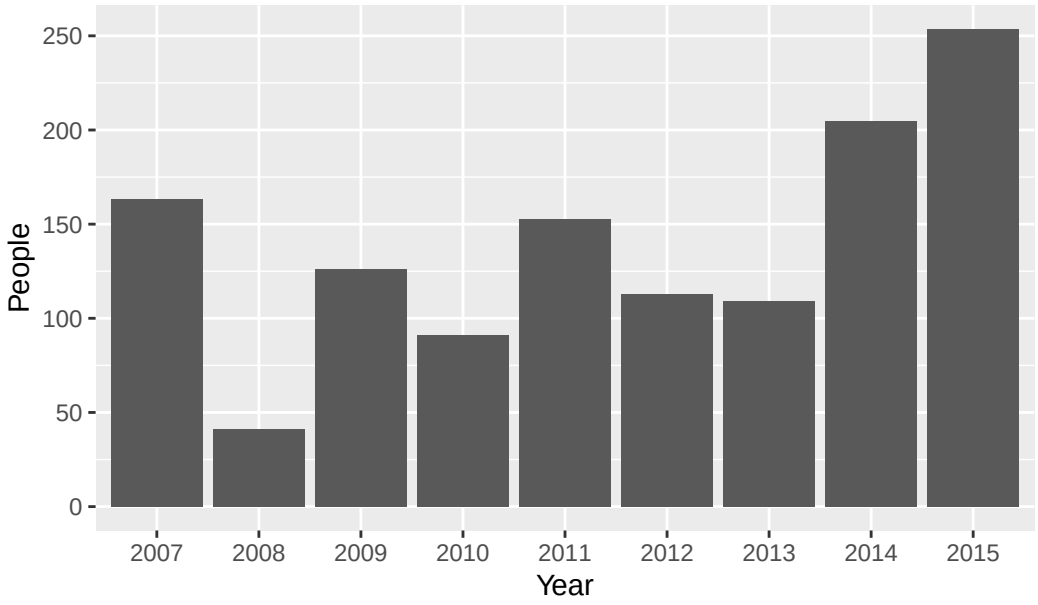

Figure 1: Average Enrollment

Premium Data

6. Merge the plan characteristics data to the dataset you created in Step 5 above. Provide a graph showing the average premium over time. Don't forget about formatting!

To do this, I merged in the market penetration and state-level data into the existing dataset created in the previous step. From here, I created a graph that demonstrated the premium price over time.

```
Warning in max(f): no non-missing arguments to max; returning -Inf
```

```
Warning: Computation failed in `stat_summary()`
Caused by error in `seq_len()`:
! argument must be coercible to non-negative integer
```

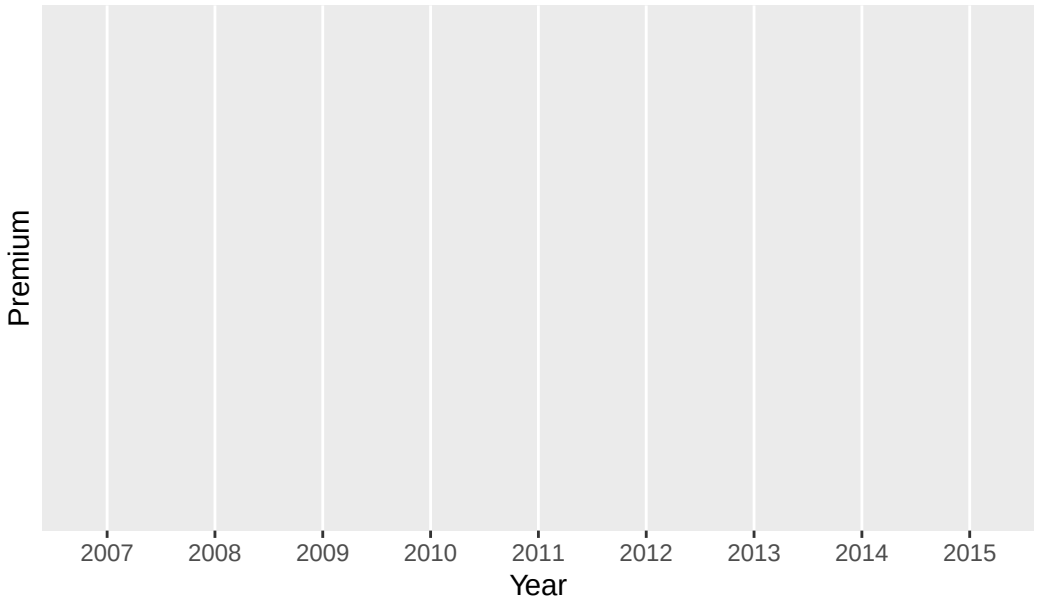

Figure 2: Average Premiums

7. Provide a graph showing the percentage of $0 premium plans over time. Also…remember to format things.

Here is the graph of the percentage of $0 premum plans over this time period:

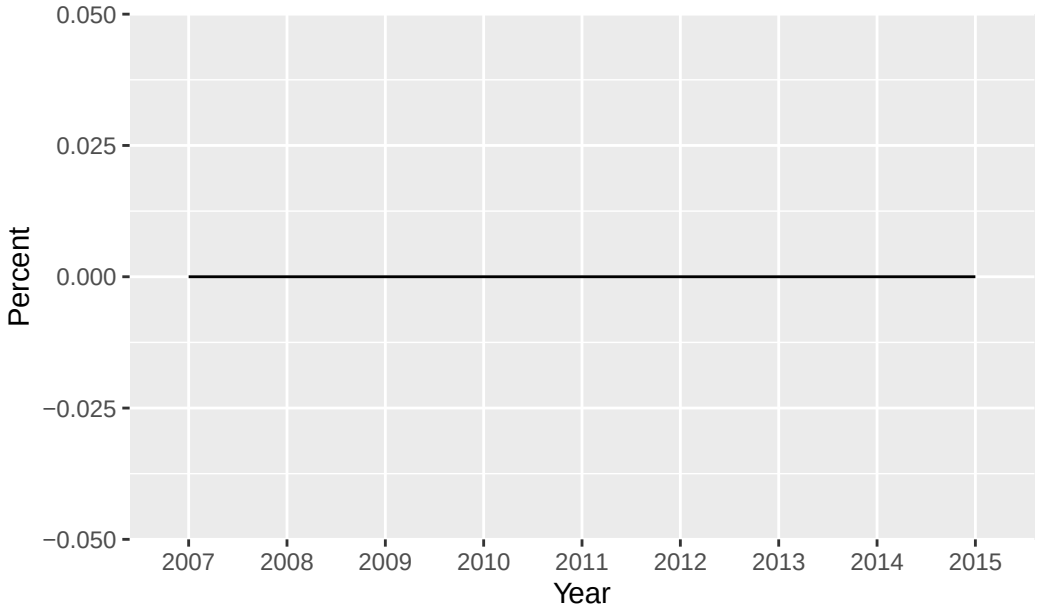

Figure 3: Share of 0 premium plans

8. Why did we drop the “800-series” plans?

800-series plans are only available to a select group of individuals. Since most people cannot opt into these types of plans, they should be excluded from the dataset.

9. Why do so many plans charge a $0 premium? What does that really mean to a beneficiary?

There is a standard premium already paid for Medicare Part B, so this $0 premium is just the excess over that standard premium. Additionally, tax dollars subsidize this program, so that helps to allow these $0 premiums to exist.

10. Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated you.

My experience working with this data involved much trial and error. With little R experience beforehand, I had to rely on my peers, online resources like git, stack overflow, and others. One thing I learned was how to review R documentation to get to the desired outcome. One thing I struggled with was using join functions and adapting my code to Quarto. But now that I have learned this, it should be smoother going forward. Additionally, my graphs did not come out correctly in Quarto, so I would like to remedy that as well.